LA-UR-19-24739
Approved for public release; distribution is unlimited.

Title: Gradient-Free Construction of Active Subspaces for Dimension Reduction

Author(s): Williams, Brian J.
Coleman, Kayla
Smith, Ralph C.
Morris, Max D.

Intended for: Archives

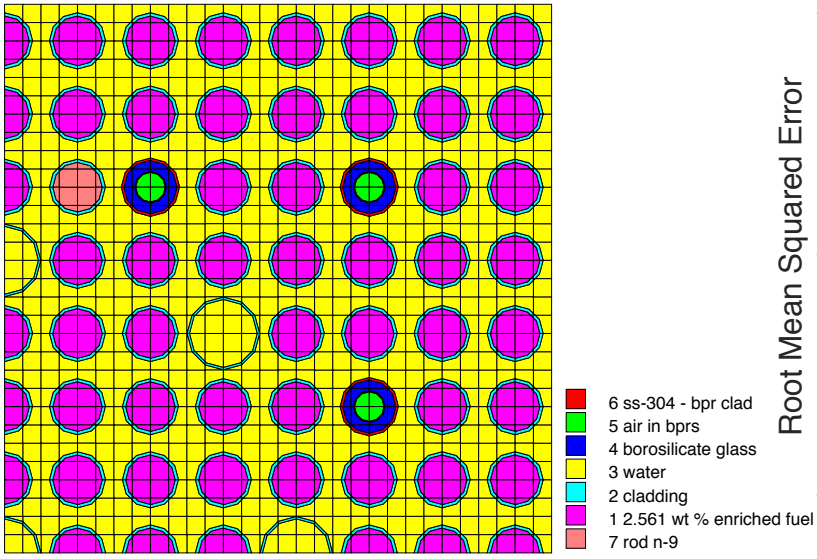
Issued: 2019-05-22

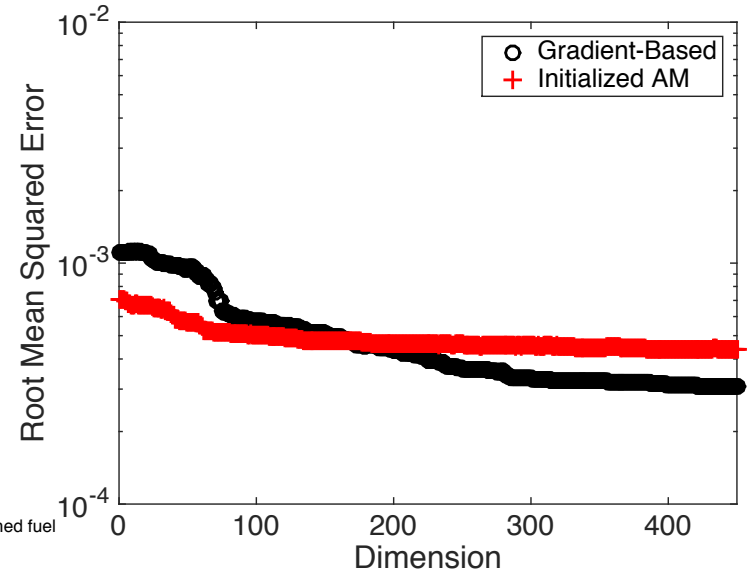
Gradient-Free Construction of Active Subspaces for Dimension Reduction

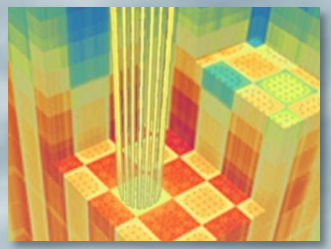
Brian J. Williams (LANL)

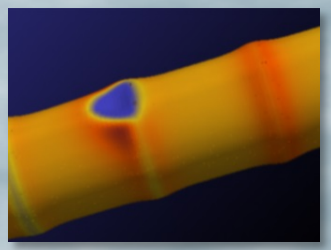
Kayla Coleman, Ralph C. Smith (NCSU)

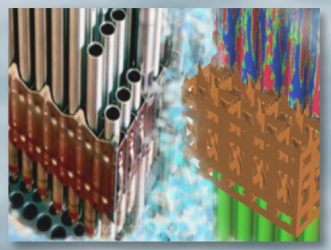
Max Morris (ISU)











The Consortium for Advanced Simulation of LWRs
A DOE Energy Innovation Hub

U.S. DEPARTMENT OF ENERGY

Abstract

Recent developments in the field of reduced order modeling - and in particular, active subspace construction - have made it possible to efficiently approximate complex models by constructing low-order response surfaces based upon a small subspace of the original high dimensional parameter space. These methods rely upon the fact that the response tends to vary more prominently in a few dominant directions defined by linear combinations of the original inputs, allowing for a rotation of the coordinate axis and a consequent transformation of the parameters. In this talk, we discuss a gradient free active subspace algorithm that is feasible for high dimensional parameter spaces where finite-difference techniques are impractical. We illustrate an initialized gradient-free active subspace algorithm for a neutronics example implemented with SCALE6.1, for input dimensions up to 7700.

Dimension Reduction

- The statistics community has been interested in dimension reduction methods for regression problems for 25+ years
  - Introduction of sliced inverse regression (SIR) and sliced average variance estimation (SAVE) in 1991
- Statistical formulation: Estimate the *central subspace*
  - Regress response Y = f(**X**) on a random *m*-vector of inputs **X**
  - Intersection of all subspaces $\mathbb{S}$ with the property that Y is conditionally independent of **X** given the projection of **X** onto $\mathbb{S}$
  - Result is a set of $n < m$ orthogonal linear combinations of **X**
- [Xia, Annals of Statistics 2007] introduced nonparametric methods to estimate $\mathbb{S}$ exhaustively
  - Compared performance to SIR, SAVE, principal Hessian direction (PHD), and minimum average variance estimation (MAVE)
- [Cook et al., JASA 2009] introduced a likelihood method for estimating $\mathbb{S}$ termed likelihood acquired directions (LAD)
  - Compared performance to SIR, SAVE, and directional regression (DR)
  - Assumes conditional normality but robust to non-normality
  - Likelihood ratio statistic, AIC, BIC used to choose *n*

Active Subspaces

Motivation:

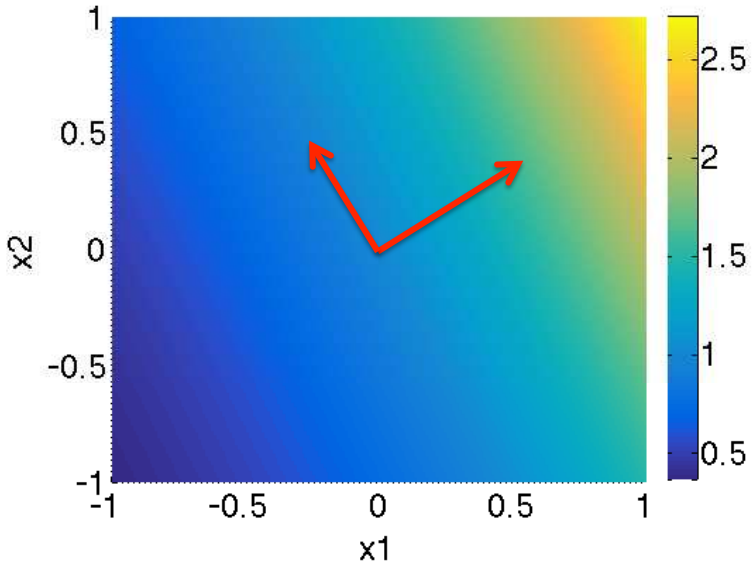
- Some UQ problems involve high-dimensional input spaces that present challenges for standard surrogate and model calibration algorithms
  - e.g. 7700 cross section perturbations in a PWR quarter fuel lattice
  - 10k – 100k+ parameters possible in CIPS Challenge Problem
- Typically sensitivity analysis would substantially reduce this dimension as most parameters have a relatively small influence on the QoIs
- Popular active subspace methods seek to find a substantially reduced set of parameters formed as *linear combinations* of the original parameters
  - Conceptual similarities to statistical dimension reduction methods
  - If possible identify a set of 100 or fewer *active* parameters
- Use gradients to identify active parameters if they are produced by the code. Otherwise, gradient free approaches must be considered
  - Active area of research

Goal: Using a new gradient free algorithm for active subspace discovery, determine active parameters for use in surrogate construction and model calibration

Active Subspace Construction

Note:

- Outputs may vary significantly in only a few “active” directions, which may be linear combinations of inputs.



Example: $y = \exp(0.7x_1 + 0.3x_2)$

- Varies most in [0.7, 0.3] direction
- No variation in orthogonal direction

Strategy:

- Employ gradient-based or gradient-free techniques, in combination with SVD or QR to construct active subspace.
- Employ active subspaces for:
  - Linear Karhunen-Loeve expansion-based UQ
  - Surrogate or reduced-order model construction
  - Model calibration

Gradient-Based Active Subspace

Active Subspace: See [Constantine, SIAM 2015]. Consider

$$f = f(\mathbf{x}),\ \mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^m$$

and

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_m} \right]^T$$

Construct outer product

$$\mathbf{C} = \int (\nabla_x f)(\nabla_x f)^T \rho dx$$

$\rho(x)$: distribution of input x parameters

Partition eigenvalues: $\mathbf{C} = \mathbf{W}\Lambda\mathbf{W}^T$

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix},\ \mathbf{W} = [\mathbf{W}_1\ \mathbf{W}_2]$$

Rotated Coordinates:

$$\mathbf{y} = \mathbf{W}_1^T \mathbf{x} \in \mathbb{R}^n \text{ and } \mathbf{z} = \mathbf{W}_2^T \mathbf{x} \in \mathbb{R}^{m-n}$$

Motivation

Results:

Derivative of f(**x**) in the direction $\mathbf{w}_i$

(1) $$\lambda_i = \int \left(\left(\nabla_{\boldsymbol{x}} f \right)^T \boldsymbol{w}_i \right)^2 \rho(\boldsymbol{x})\, d\boldsymbol{x}$$

(2) $$\int \left(\nabla_{\mathbf{z}} f \right)^T \left(\nabla_{\mathbf{z}} f \right) \rho(\mathbf{x})\, d\mathbf{x} = \lambda_{n+1} + \cdots + \lambda_m$$

- *n* can be chosen by looking for a ”large” gap between λ_n and λ_{n+1}, such that $\lambda_{n+1} + \ldots + \lambda_m$ is relatively “small”

(3) $$f(\boldsymbol{x}) \approx g(\mathbf{W}_1^T \mathbf{x})$$ g is a *link function*

Active and Central Subspaces

- Suppose f(**x**) = g(**y**) for $\boldsymbol{y} = \boldsymbol{W}_1^T \boldsymbol{x}$

$$\begin{aligned}\pi(f(\boldsymbol{x}), \boldsymbol{x}|\boldsymbol{y}) &= \pi(g(\boldsymbol{y}), \boldsymbol{x}|\boldsymbol{y}) \\ &= \pi(g(\boldsymbol{y})|\boldsymbol{y}, \boldsymbol{x})\, \pi(\boldsymbol{x}|\boldsymbol{y}) \\ &= \pi(g(\boldsymbol{y})|\boldsymbol{y})\, \pi(\boldsymbol{x}|\boldsymbol{y}) \\ &= \pi(f(\boldsymbol{x})|\boldsymbol{y})\, \pi(\boldsymbol{x}|\boldsymbol{y})\end{aligned}$$

- Inputs and output are therefore conditionally independent given the active variables, and so the active subspace defined by the columns of $\mathbf{W}_1$ contains the central subspace

Estimation

Approximation via Monte Carlo:

1. Draw *M* samples $\{ \mathbf{x}_j \}$ independently from $\rho(\mathbf{x})$
2. For each $\mathbf{x}_j$, compute $\nabla_{\mathbf{x}} f_j = \nabla_{\mathbf{x}} f(\mathbf{x}_j)$
3. Approximate

$$\mathbf{C} \approx \hat{\mathbf{C}} = \frac{1}{M} \sum_{j=1}^{M} \left(\nabla_{\mathbf{x}} f_j\right) \left(\nabla_{\mathbf{x}} f_j\right)^T$$

4. Compute the eigendecomposition $\hat{\mathbf{C}} = \hat{\mathbf{W}} \hat{\Lambda} \hat{\mathbf{W}}^T$

Steps 3 and 4 equivalent to computing the SVD of the *gradient matrix*

$$\mathbf{G} = \frac{1}{\sqrt{M}} \left[\nabla_{\mathbf{x}} f_1 \cdots \nabla_{\mathbf{x}} f_M\right] = \hat{\mathbf{W}} \hat{\Lambda}^{1/2} \hat{\mathbf{V}}$$

Error in estimated active subspace:

$$\varepsilon = ||\boldsymbol{W}_1 \boldsymbol{W}_1^T - \hat{\boldsymbol{W}}_1 \hat{\boldsymbol{W}}_1^T||_2 = ||\hat{\boldsymbol{W}}_1^T \boldsymbol{W}_2||_2$$

$$\varepsilon \leq \frac{4\lambda_1 \delta}{\lambda_n - \lambda_{n+1}}$$

δ is a user-specified tolerance for the eigenvalue estimates (used to choose M)

Order Determination

1. Gap-based
   - Stop at largest gap in eigenvalue spectrum
2. Error-based
   - Specify error tolerance ε_{tol}, $\mathbf{G} = \mathbf{U}\,\Lambda^{1/2}\,\mathbf{V}^T$

   a) Draw a sequence of p standard Gaussian vectors $\{\omega_1, \ldots, \omega_p\}$

   b) Let $\tilde{\mathbf{U}}_{m\times j}$ be the first j columns of $\mathbf{U}$

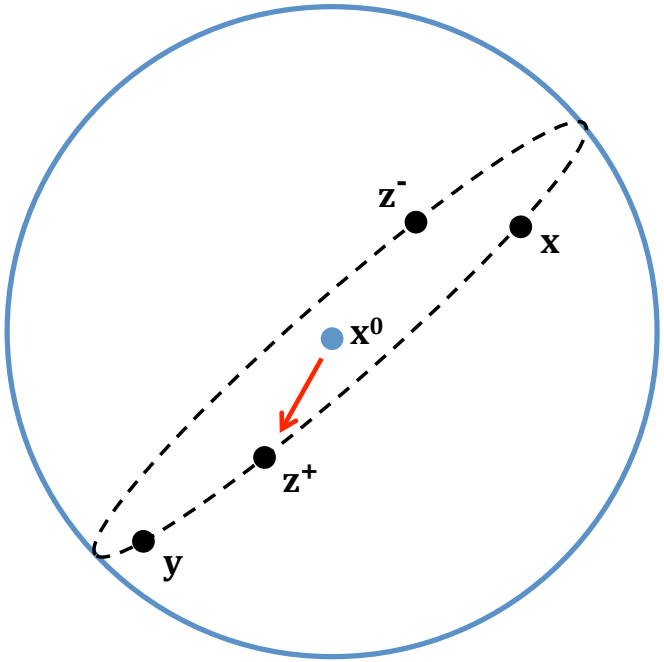
   c) Let $\varepsilon_{\text{upp}}^{j} = 10\sqrt{2/\pi}\,\max_{i=1,\ldots,p} ||(\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T)\mathbf{G}\omega^i||$

   - Find smallest j for which $\varepsilon_{\text{upp}}^{j} < \varepsilon_{\text{tol}}$
   - Error bound holds with probability $1 - 10^{-p}$
3. PCA-based
   - Stop at minimal dimension yielding variance explained in covariance matrix formed from **G** exceeding user-specified threshold (e.g. 99%)
4. Response surface-based
   - Use the minimal dimension required to reduce response surface error on a validation dataset below a user-specified threshold (e.g. 0.01, 0.001)

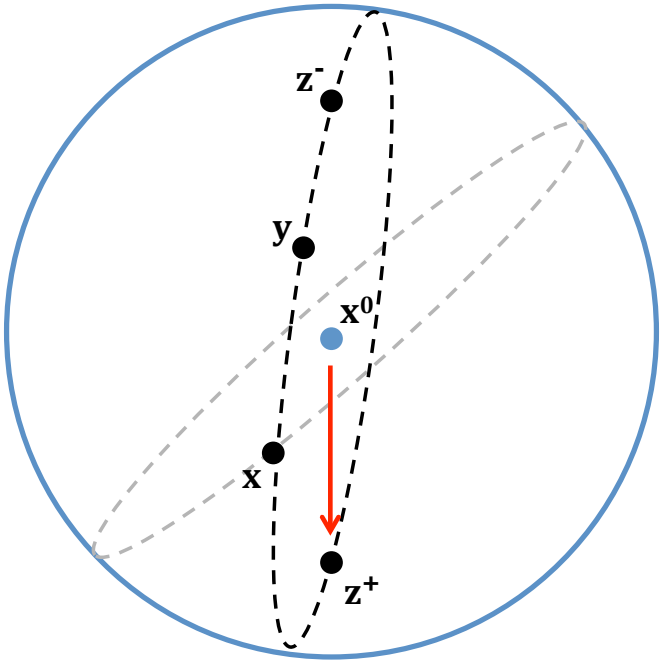
Goal: Determine dimension of active subspace

Gradient Approximation for Large Input Spaces

- Utilized when finite difference approach to gradient approximation is computationally prohibitive; e.g., SCALE6.1 with 7700 inputs.
- Construct ellipsoid where linearity is reasonable assumption.
- Maximize function values and gradient information using “great ellipsoid” relations.



Iteration 1



Iteration 2

"Great Ellipsoid" Solution

- Consider a matrix **C** collecting *h*+1 input samples from the surface of the unit hypersphere:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{w} & \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_h \end{bmatrix}$$

- Collect the sampled output differences into a vector **y**:

$$\boldsymbol{y} = \begin{bmatrix} g(\boldsymbol{w}) - g(\boldsymbol{0}) & g(\boldsymbol{v}_1) - g(\boldsymbol{0}) & \cdots & g(\boldsymbol{v}_h) - g(\boldsymbol{0}) \end{bmatrix}^T$$

- The direction of steepest ascent within the column space of **C** is given by:

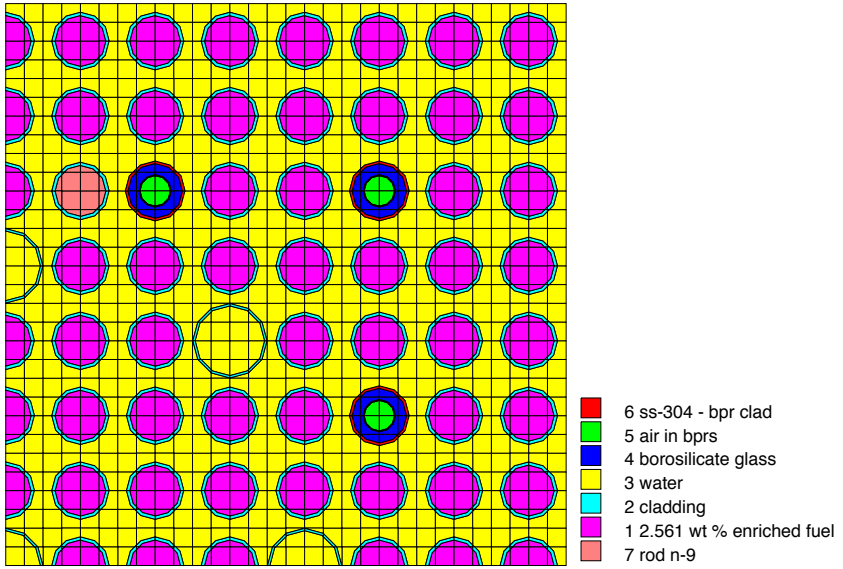
$$\boldsymbol{u}_{\max} = \frac{\boldsymbol{C}\left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-}\boldsymbol{y}}{\sqrt{\boldsymbol{y}^T\left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-}\boldsymbol{y}}}$$

SCALE6.1: High-Dimensional Example

Setup:

- Input Dimension: 7700
- Output k_{eff}

Materials			Reactions	
$^{234}_{92}\mathrm{U}$	$^{10}_{5}\mathrm{B}$	$^{31}_{15}\mathrm{P}$	Σ_t	$n \rightarrow \gamma$
$^{235}_{92}\mathrm{U}$	$^{11}_{5}\mathrm{B}$	$^{55}_{25}\mathrm{Mn}$	Σ_e	$n \rightarrow p$
$^{236}_{92}\mathrm{U}$	$^{14}_{7}\mathrm{N}$	$_{26}\mathrm{Fe}$	Σ_f	$n \rightarrow d$
$^{238}_{92}\mathrm{U}$	$^{15}_{7}\mathrm{N}$	$^{116}_{50}\mathrm{Sn}$	Σ_c	$n \rightarrow t$
$^{1}_{1}\mathrm{H}$	$^{23}_{11}\mathrm{Na}$	$^{120}_{50}\mathrm{Sn}$	$\bar{\nu}$	$n \rightarrow {}^{3}\mathrm{He}$
$^{16}_{8}\mathrm{O}$	$^{27}_{13}\mathrm{Al}$	$_{40}\mathrm{Zr}$	χ	$n \rightarrow \alpha$
$_{6}\mathrm{C}$	$_{14}\mathrm{Si}$	$_{19}\mathrm{K}$	$n \rightarrow n'$	$n \rightarrow 2n$



PWR Quarter Fuel Lattice

Note: We cannot efficiently approximate all directional derivatives required to approximate the gradient matrix. Requires an efficient gradient approximation algorithm.

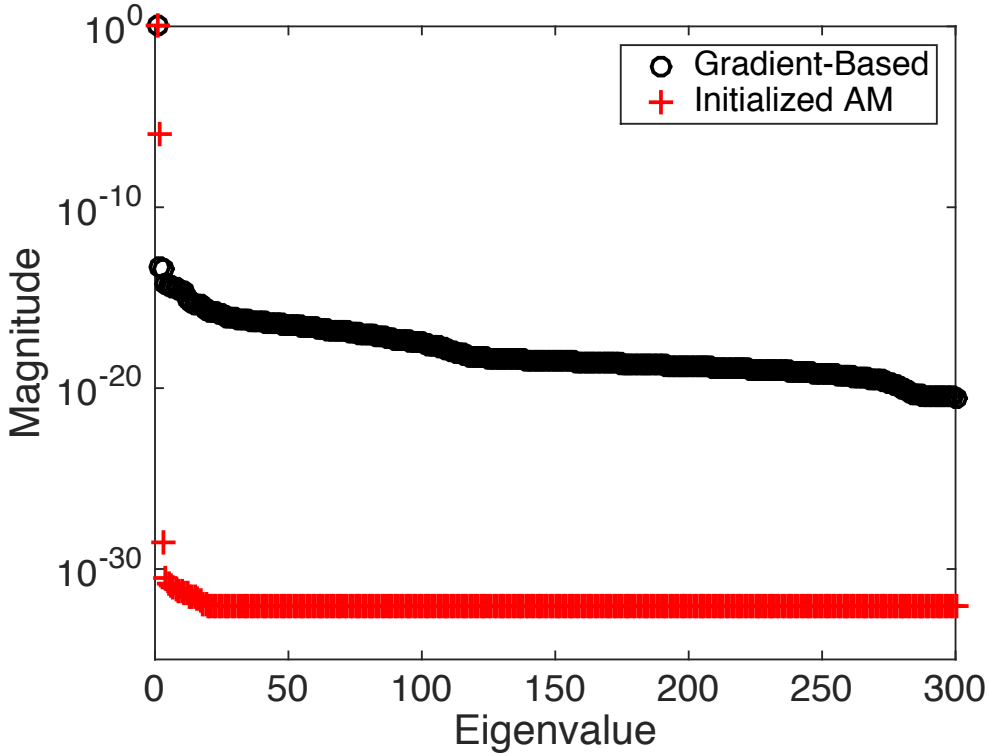
SCALE6.1: High-Dimensional Example

Setup:

- Input Dimension: 7700

SCALE Evaluations:

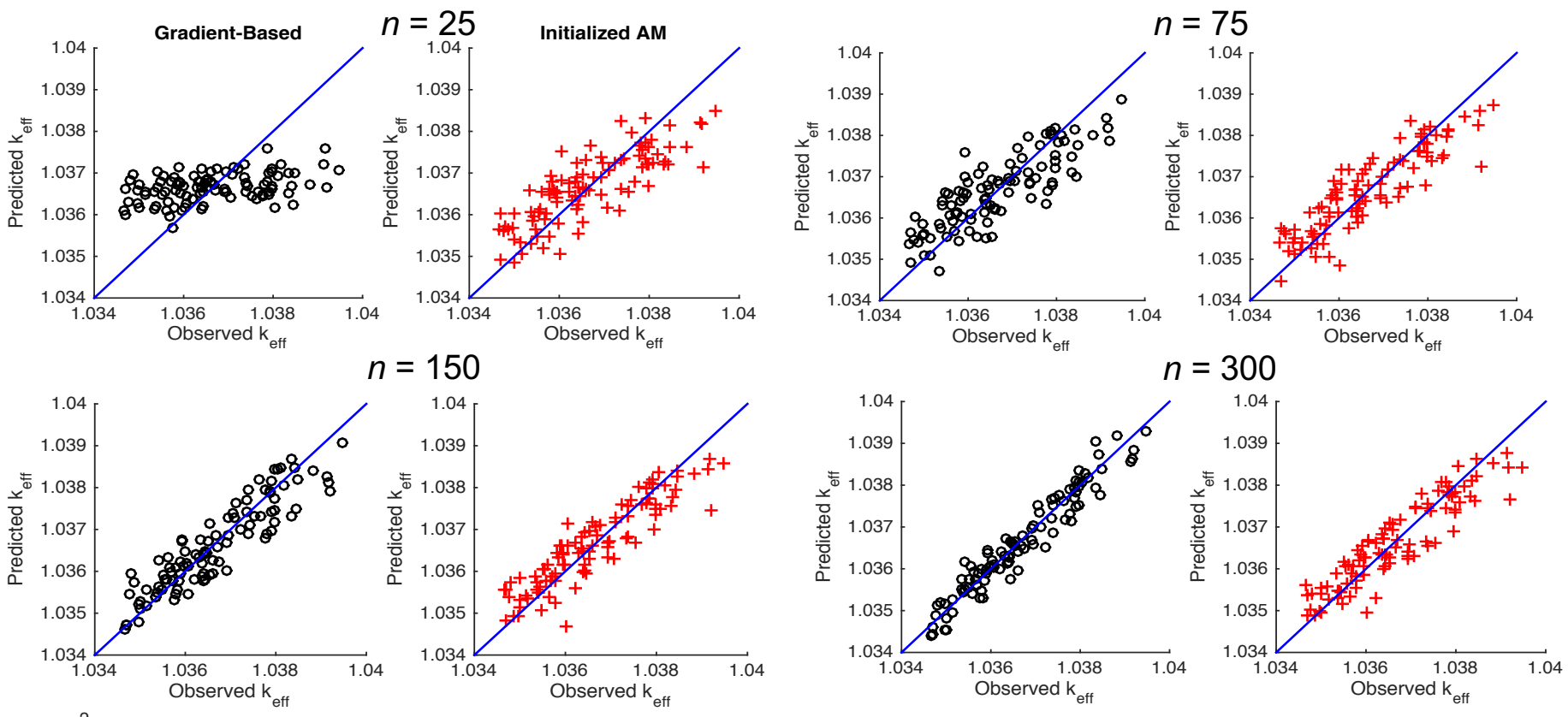
- Gradient-Based: 1000
- Initialized Adaptive Morris: 18,392 (0.20%)
- Projected Finite-Difference: 7,701,000



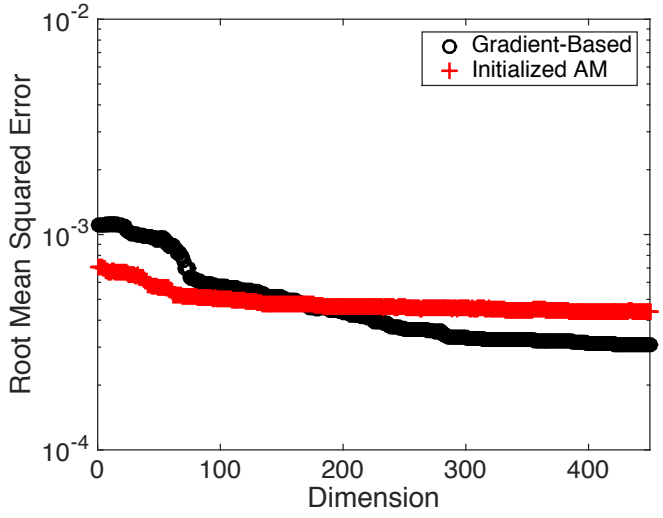
Active Subspace Dimensions:

	Gap	PCA				Error Tolerance			
Method		0.75	0.90	0.95	0.99	10^{-3}	10^{-4}	10^{-5}	10^{-6}
Gradient-Based	1	2	6	9	24	1	13	90	233
Initialized AM	1	1	1	1	2	1	2	2	2

SCALE6.1: High-Dimensional Example



Predicted vs. Observed
Gradient-Based and Initialized AM



Response surface error (RMSE) vs.
Active Subspace dimension (*n*)

Improved Gradient Approximation

- Can the function evaluations utilized for gradient approximation be selected more efficiently?
- At iteration *i*, the direction of steepest ascent within a randomly determined subspace M_i (which also contains the direction of steepest ascent from iteration $i - 1$ for $i > 1$) is determined
- For the assumed linear approximation, at iteration *i* the function does not vary in the orthocomplement O_i in M_i of the direction of steepest ascent
- At iteration *i*, define a subspace S_i spanned by the accumulated orthocomplements from previous iterations ($S_i = \text{span}\{O_1, \ldots, O_{i-1}\}$), and ensure the subspace M_i in which the steepest ascent direction is to be found is restricted to the orthocomplement of S_i
- At most *d* iterations required to converge to the gradient:

$$\sum_{i=1}^{d} \dim(M_i) = m + d - 1$$

Quality of Gradient Approximation

- Consider a *k*-dimensional subspace defined by the column space of a matrix **M** in which the gradient is currently approximated by $\mathbf{z}^+$. It can be shown that

$$\boldsymbol{z}^+ = \frac{\boldsymbol{P_M}\left(\nabla_{\boldsymbol{x}} f\right)}{||\boldsymbol{P_M}\left(\nabla_{\boldsymbol{x}} f\right)||}$$

- We assume the unknown normalized gradient vector **z** is uniformly distributed on the unit sphere, and consider the distribution of the cosine of the angle between the random quantities **z** and $\mathbf{z}^+$:

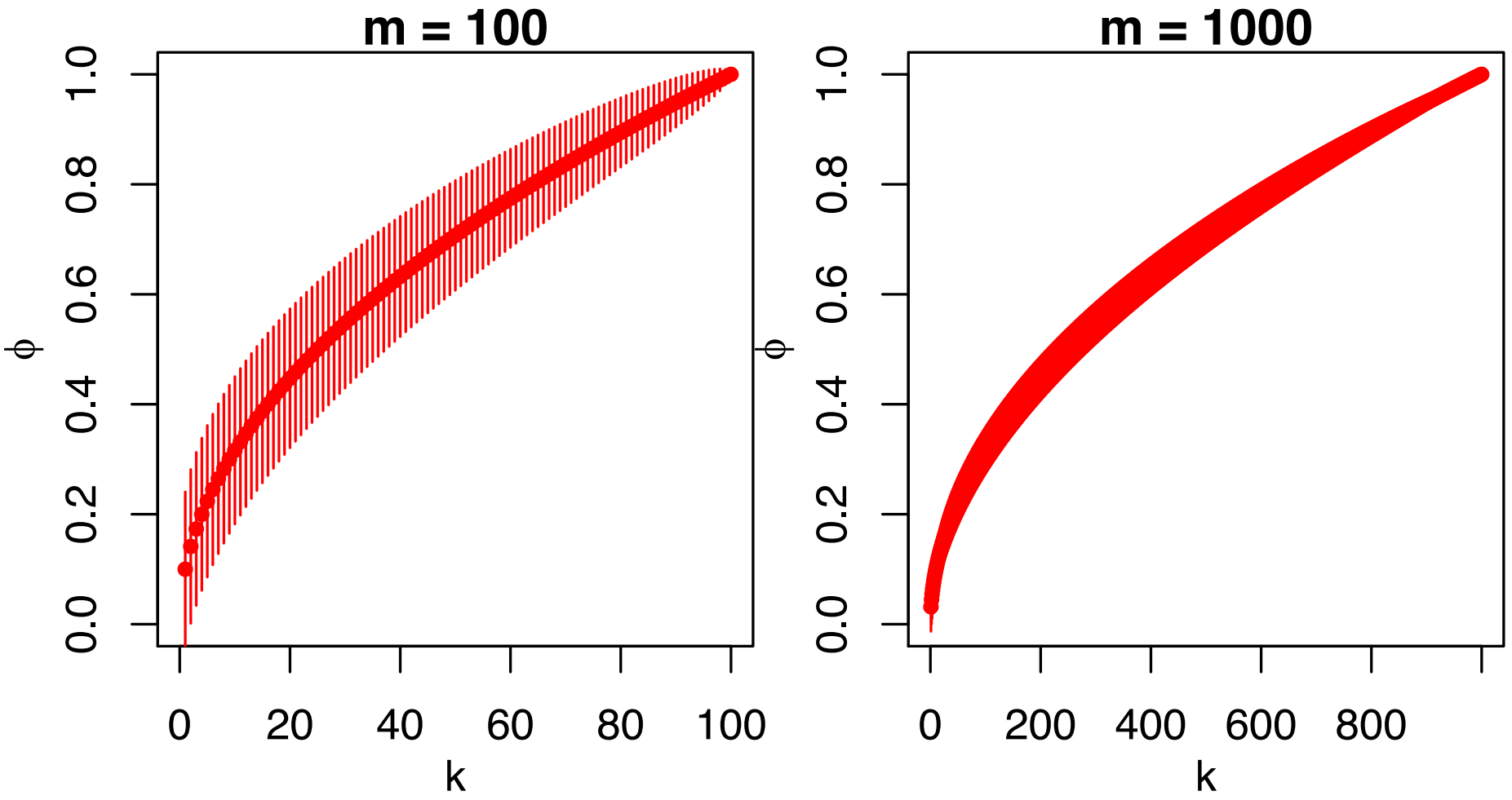
$$\phi = \sqrt{\frac{\boldsymbol{z}^T \boldsymbol{P_M} \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{z}}}\,,\ \boldsymbol{z} \sim N_m(\boldsymbol{0}, \boldsymbol{I}_m)$$

- The mean and standard deviation of ϕ are approximated as follows:

$$E[\phi] \approx \sqrt{\frac{k}{m}}\,,\ SD[\phi] \approx \frac{1}{m}\sqrt{\frac{m-k}{2}}$$

Quality of Gradient Approximation

- Uncertainty in error decreases with increasing input dimension



Elliptic PDE: Moderate-Dimensional Example

- Consider the following equation:

$$-\nabla_{\boldsymbol{s}} \cdot (a(\boldsymbol{s}, \boldsymbol{x}) \nabla_{\boldsymbol{s}} u(\boldsymbol{s}, a(\boldsymbol{s}, \boldsymbol{x}))) = 1\,,\ \boldsymbol{s} \in [0,1]^2$$

- Boundary conditions: $u = 0$ (left, top, bottom); $\dfrac{\partial u}{\partial s_1} = 0$ on right (Γ_2)

- $a(\mathbf{s}, \mathbf{x})$ is taken to be a log-Gaussian second-order random field (m = 100):

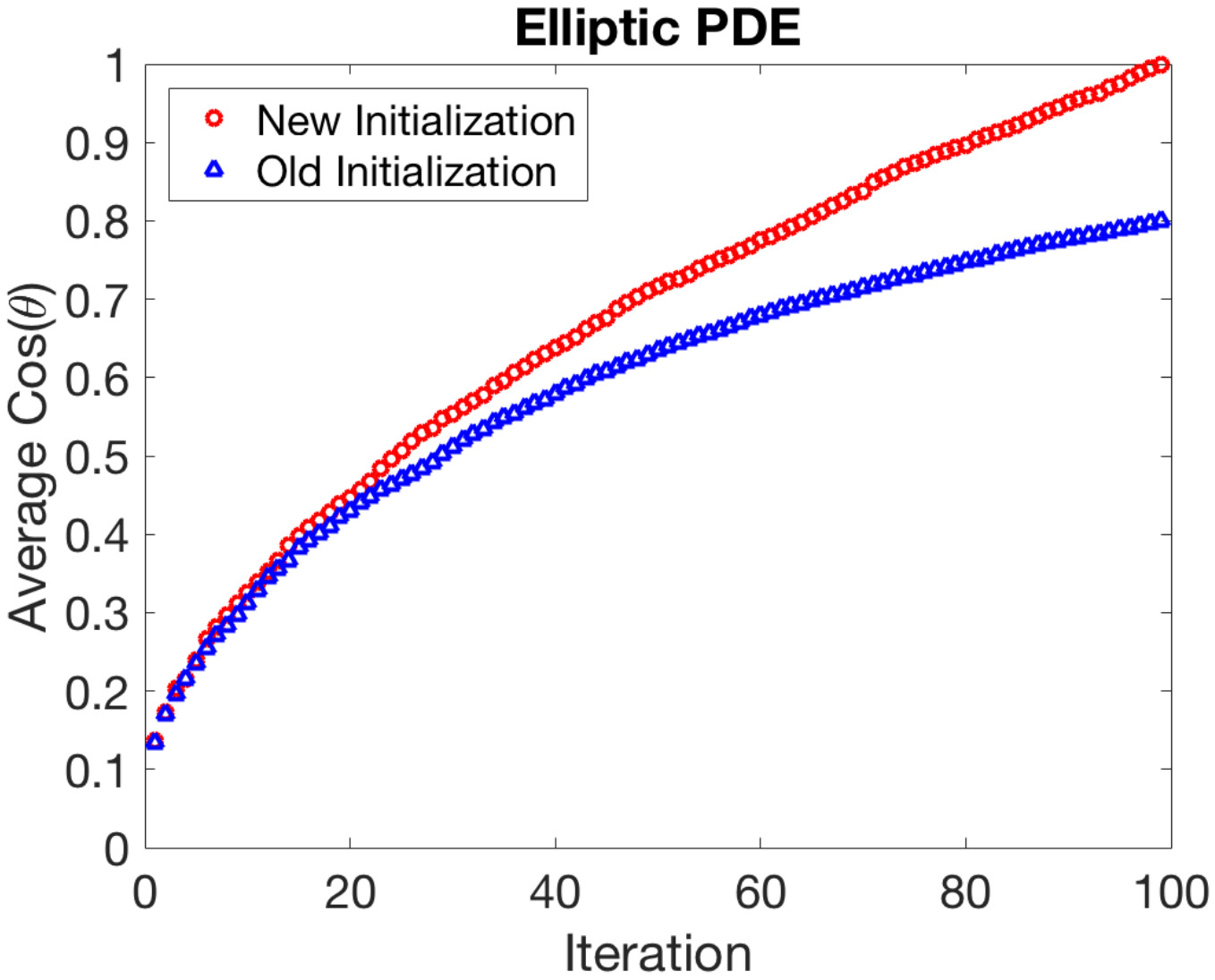
$$\log(a(\boldsymbol{s}, \boldsymbol{x})) = \sum_{i=1}^{m} x_i \sqrt{\gamma_i}\, \phi_i(\boldsymbol{s})$$

- Response of interest:

$$f(\boldsymbol{x}) = \frac{1}{|\Gamma_2|} \int_{\Gamma_2} u(\boldsymbol{s}, \boldsymbol{x})\, d\boldsymbol{s}$$

- Standard finite element method used to discretize this elliptic problem, producing $f(\mathbf{x})$ and the adjoint-computed $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$

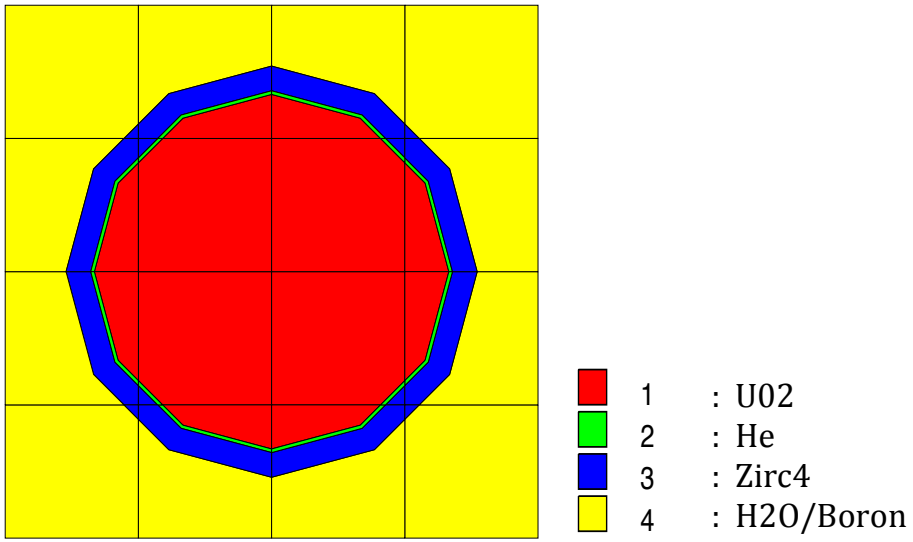
Elliptic PDE: Moderate-Dimensional Example



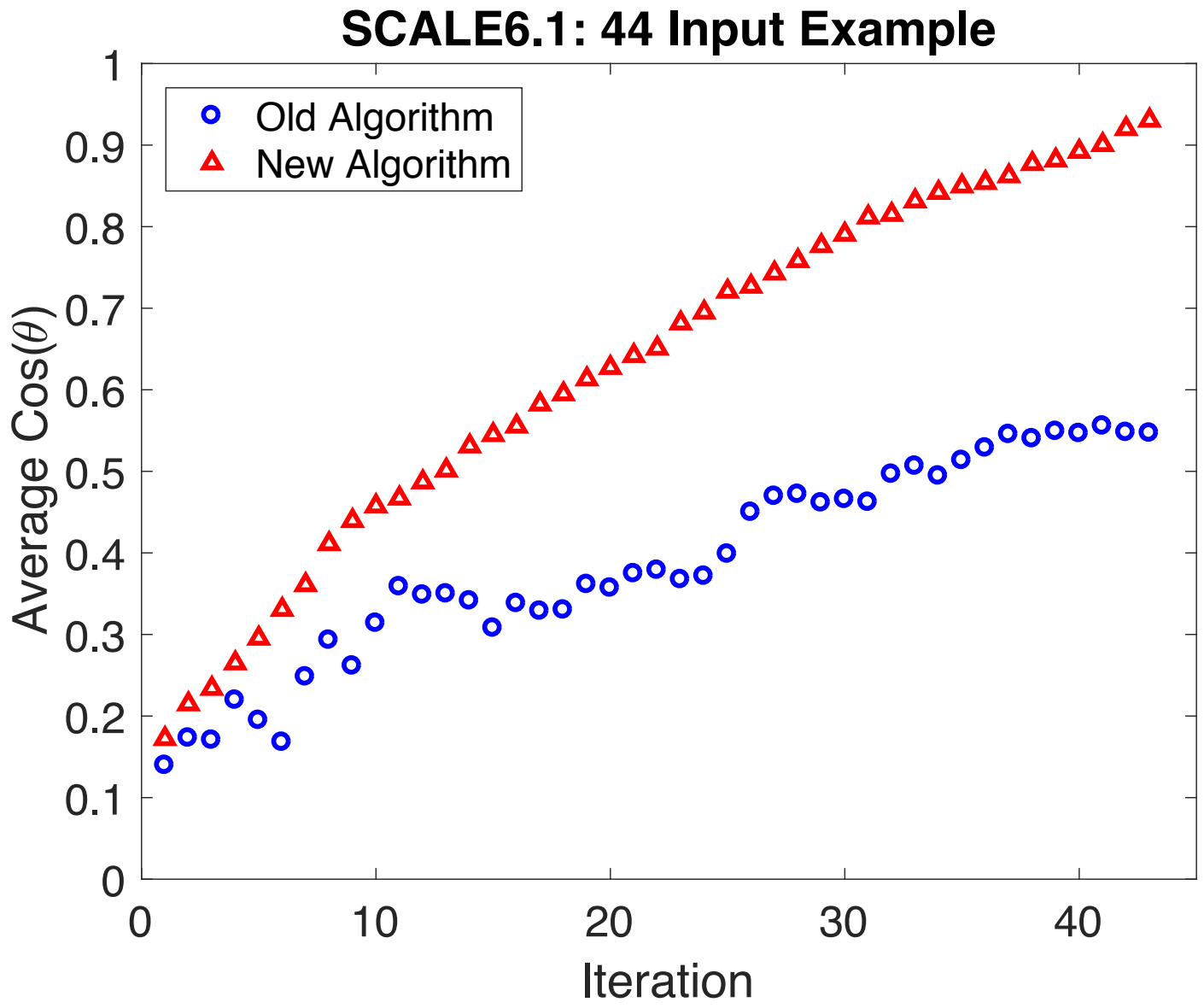
SCALE6.1: Moderate-Dimensional Example

Setup:

- Material: U_{92}^{235}
- Cross-sections: $\Sigma_f(E)$
- Energy groups: 44
- Total input dimension: 44
- Output: k_{eff}



SCALE6.1: Moderate-Dimensional Example



Gradient-Free Active Subspaces

Observations:

- If available, use gradient information to identify active subspaces.
- Many legacy codes do not calculate gradients. In these cases, gradient-free active subspace discovery is required.
- For complex codes, strategies required to reduce computational effort.

Papers:

- A. Lewis, R.C. Smith and B. Williams (2016), "Gradient free active subspace construction using Morris screening elementary effects," *Computers and Mathematics with Applications,* 72(6), 1603-1615.
- K.D. Coleman, A. Lewis, R.C. Smith, B. Williams, M. Morris and B. Khuwaileh (2019), "Gradient-free construction of active subspaces for dimension reduction in complex models with applications to neutronics," *SIAM/ASA Journal on Uncertainty Quantification*, 7(1), 117-142.

Present and Future Work:

- Integrate gradient approximation algorithm into Sandia's Dakota software.
- Continued investigation of response surfaces constructed from active parameters in Bayesian model calibration applications.